The Phenomenological Critics of Folk Psychology: The Case of False Belief

Mitchell Herschbach

Department of Philosophy – 0119
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0119
mherschb@ucsd.edu

Abstract:
The dominant account of human social understanding is that we possess a “folk psychology”: that we can understand and interact with other people because we have a commonsense understanding of their mental states. However, recently a group of philosophers from the phenomenological tradition have called into question the folk psychological account of human social understanding. I will first explicate the “phenomenological critics’” account of social understanding, making use of Michael Wheeler’s distinction between online and offline intelligence. I will then discuss recent studies of early false belief understanding, and identify ways it supports and ways it challenges the critics’ account.

1. Social Understanding and “Folk Psychology”

Over the last few decades, philosophical and empirical discussions of human social understanding have been formulated in terms of “folk” or “commonsense” psychology: that it is our folk understanding of human psychology which explains our abilities to understand and interact with other people (e.g., Carruthers & Smith, 1996; Davies & Stone, 1995; Goldman, 2006). Talk of folk psychology (FP) is so commonplace now that it may seem a necessary catchall term for whatever knowledge of human psychology and behavior most humans possess that makes possible human social interaction. But FP as usually characterized is not so theoretically uncommitted, instead involving specific assumptions about the nature of
interpersonal understanding, and the role of such understanding in social interaction. As Ratcliffe and Hutto (in press) summarize:

the received wisdom about folk psychology encapsulates two chief assumptions: (i) that making sense of actions requires interpreting them in terms of reasons composed of various propositional attitudes (at a bare minimum – beliefs and desires) and (ii) that this activity is primarily concerned with providing predictions and explanations of actions.

(Introduction)

The two major accounts of FP, theory theory (TT) and simulation theory (ST), share these basic assumptions, treating FP as involving the attribution of propositional attitudes to others in order to explain and predict their behavior. All forms of TT and most forms of ST treat such propositional attitude ascription to require the possession of various mental state concepts, e.g., BELIEF and DESIRE, according to which mental states are conceived of as the inner causes of behavior. Much ink has been spilled distinguishing TT and ST, with many researchers moving towards hybrid theory-simulation accounts, taking both theory and simulation to have necessary roles. For most philosophers and scientists, TT, ST, or some combination of the two, are the only theoretical options and the assumptions they share go unquestioned. FP is presumed to be the proper way to frame any investigation of the nature of human social understanding, whether the behavioral experiments of developmental psychologists, neuroimaging studies of neuroscientists, or the theoretical work of philosophers. It is our proficiency with FP that is offered as an explanation for human social interaction—that without FP, people’s navigation of the social world would be significantly impaired.

These assumptions of FP can be made more concrete by examining a dominant experimental paradigm for studying social understanding: the false belief task. A common version used with children is the Sally-Anne task, where dolls (named Sally and Anne) are often used in place of actual humans as protagonists (Baron-Cohen, Leslie, & Frith, 1985). The child-
subject observes Sally put a marble in a basket then leave the room. While Sally is away, Anne enters and moves the marble from the basket into a box also in the room. Sally then reenters, and the child is asked, “Where will Sally look for the marble?” The child knows that the marble is currently in the box. The child must suppress this information, and consider where Sally thinks the marble is located. The child must remember that Sally last saw the marble in the basket and that she is unaware of Anne’s having moved the marble. Thus the correct response to the experimenter’s question is that Sally believes (falsely) that the marble is in the basket.

Passing the false belief task is taken as signaling a significant development in the acquisition of a FP-understanding of other people. In particular, such understanding requires the child to possess a concept of belief and thereby understand the representational nature of belief, i.e., that people’s beliefs can fail to accord with reality. Having a representational understanding of mental states is crucial to FP’s conception of social understanding, as it allows one to understand people’s actions when they are based on inaccurate representations of reality. Having such an understanding can expand the forms of social interaction available to an agent. For instance, it permits the manipulation of other’s representational mental states in order to deceive them, which can be important in competitive situations. While theory-theorists and simulation theorists disagree about what exactly explains the ability to pass the false belief task, both take such tasks to be paradigmatic of our FP-social understanding, and success in this domain to be the result of mental state attributions.

Recently, researchers have called into question the FP-account of social understanding. In the next section I will explicate criticisms of FP offered by philosophers working in the phenomenological tradition. After spelling out this alternative conception of social
understanding, I will address new empirical research on false belief understanding, and the challenge it poses to this alternative framework.

2. **The Phenomenological Critics of Folk Psychology**

Shaun Gallagher, Matthew Ratcliffe, Daniel Hutto, Dan Zahavi, and others working in the phenomenological tradition have offered critiques of ST and TT, and FP in general, and proposed alternative accounts of the nature of human social understanding (e.g., Gallagher, 2001, 2005, in press; Gallagher & Hutto, in press; Ratcliffe, 2006a, 2006b; Zahavi, 2005). These phenomenological critics of FP, as I will call them, claim that our social understanding does not primarily involve propositional attitude ascription for the purpose of explanation and prediction. The nature of their attack on FP has often been characterized with regard to the standard false belief task (Gallagher, 2001, 2005; Ratcliffe, 2006b, ch. 4).

FP-accounts take the child’s situation in the false belief task as paradigmatic: it requires the child to observe someone’s behavior, and predict her behavior by attributing propositional attitudes to her, in this case a false belief. If shown Sally unsuccessfully looking for the marble in the basket, the child would be expected to explain this behavior by saying that Sally wanted the marble and falsely believed that it was in the basket. The phenomenological critics point out that the false belief task places the child in the role of theorist, providing conscious explanations and predictions of a third party’s behavior through applications of the concept belief. Even if the child is not required to provide verbal explanations and predictions, the child is at least required to somehow report to the experimenter a behavioral prediction, perhaps by pointing to a location. The two dominant FP accounts, TT and ST, attempt to explain precisely this explicit explanation and prediction of behavior based on mental state attribution.
After thus characterizing the nature of the child’s cognitive stance in the false belief task, the phenomenological critics then propose the following criticism: what is not being tested, nor being explained by TT and ST, about the child’s behavior in the false belief task is the child’s interaction with the experimenter. Even if a child is too young to pass a false belief task, she is still able to talk to the experimenter, understand the experimenter’s instructions to observe the scene involving Sally and Anne, and attempt to answer the experimenter’s question about Sally’s behavior. The false belief task does not investigate the socio-cognitive abilities necessary for such social interaction, where the child is a participant in the interaction. The FP-account of social understanding, with its focus on the reflective/theoretical capacities of explaining and predicting behavior, does not obviously apply to such participatory social interactions, even though most FP-account make this assumption. While it is possible that the theorizing and simulation play a role in such social interactions, tasks placing the subject in the role of theorizer rather than participant will not be of much help in determining if this is the case. The critics further suggest that it is not obvious that the psychological processes involved in participatory social understanding are best characterized as explanation and prediction.

This is a central part of the phenomenological critique of FP as an account of social understanding: the social phenomena for which TT and ST are most clearly explanations are only a subset of the kinds of social interactions we humans engage in; and it is questionable whether the FP-accounts developed to explain reflective/theoretical social understanding can be extended to explain instances of participatory social understanding. Behavioral experimental paradigms such as the standard false belief task, and neuroimaging experiments involving reflective tasks—e.g., reading stories involving people exhibiting various mental states (see, e.g., H. L. Gallagher et al., 2000)—that are claimed to provide empirical data regarding our psychological
understanding of others, are often limited to such reflective phenomena and do not necessarily speak to the understanding required for other forms of social interaction.

The distinction between theoretical and participatory social understanding which the phenomenological critics want to call attention to can be helpfully understood in terms of Wheeler’s (2005) distinction between “online” and “offline” intelligence (which he appeals to in developing a Heideggerian framework for cognitive science). Online intelligence involves an organism’s active sensorimotor engagement with the world: “A creature displays online intelligence just when it produces a suite of fluid and flexible real-time adaptive responses to incoming sensory stimuli” (p. 12). Offline intelligence, in contrast, is exhibited when an organism is not acting, but rather reflecting on the world and what it should do. This is not a contrast between psychological processes that are explicit and available to consciousness, and ones that are not. Rather, it is about the stance an organism takes toward its environment: online sensorimotor interaction versus disengaged contemplation.

While the standard false belief task requires the child to provide a prediction or explanation to the experimenter, and thus involves interaction, FP clearly focuses more on offline forms of social understanding, where we are thinking about other people’s behavior and making explicit judgments about their mental states—in the case of the false belief task, making use of the concept belief to think about the false belief of a third party in order to explain or predict behavior. What the phenomenological critics want to call our attention to are online forms of social understanding, such as the child’s active engagement with the experimenter. Gallagher (2005) provides the following list of the sorts of online phenomena the phenomenologists have in mind: “imitation, intentionality detection, eye-tracking, the perception of meaning and emotion in movement and posture, and the understanding of intentional or goal-directed movements in
pragmatic contexts” (p. 230). He calls these “embodied practices—practices that are emotional, sensory-motor, perceptual, and non-conceptual” (p. 224). Gallagher’s language here is representative of the phenomenological critics’ proposal that these capacities for online social interaction are not amendable to description in the languages of TT or ST. In other words, the critics suggest that while there are some circumstances in which we make (offline) mental state ascriptions for the purpose of explanation and prediction via theorizing or simulation, the “embodied practices” constituting online social understanding do not involve mental state attribution, nor explanation and prediction.

These phenomenologists are not alone in recognizing the disconnect between online social understanding and the standard experimental paradigms using offline tasks. For example, Slaughter and Repacholi (2003), in their recent introduction to a volume on individual differences in “theory of mind” (i.e., FP), suggest that there are “Several dimensions of difference between laboratory theory of mind tasks and everyday social reasoning. … For instance, in everyday mind reading, we compute mental states online, and often act on these computations. It would seem rare for us to explicitly reflect on the mental state attributions we make in the course of social interactions; instead, we are much more likely to act on those attributions with an immediate behavioral or linguistic response” (p. 7, italics added). This view is also beginning to make an impact in the neuroimaging literature, where some researchers have begun using online rather than offline tasks of social understanding. These include tasks involving social perception—e.g., perception of people’s actions and their expressions of emotions and pain (see, e.g., Decety & Grèzes, 2006)—and approximations of or even actual social interaction (Gallagher & Frith, 2003; McCabe et al., 2001; Oberman et al., in press; Schilbach et al., 2006; Spiers & Maguire, 2006). While such research is often still cast in the
language of FP, the critics are clearly not alone in emphasizing online forms of social understanding.

3. **Online False-Belief Understanding**

In highlighting the distinction between what I’m calling online and offline forms of social understanding, the phenomenological critics have clearly identified a relevant distinction to which researchers must pay attention, and increasingly are doing so. But their use of the standard false belief task to illustrate their objections to the FP account of social understanding is misleading. Rather than merely marking the online–offline distinction, in their exposition they switch the type of information about the other person, or the types of properties attributed to that person, being considered. Specifically, they switch from false belief understanding (which the standard false belief task intends to test, at least in its offline form), to understanding other properties of persons relevant to communicative interaction, e.g., their intentions, emotions, and attentional focus. But they say little about false beliefs. The critics clearly want: (1) to deemphasize the significance placed of the standard false belief task, since it only focuses on offline forms of social understanding; and (2) to place greater importance on forms of online social understanding. But they also claim that online social intelligence should not be interpreted in terms of TT and ST, i.e., as not involving the application of mental state concepts for the purpose of mental state attribution. Since the critics’ discussion of the standard false belief task does not directly address their views about belief, and instead seems to change the topic, it is reasonable to ask: what exactly is their view on false-belief understanding? Is the FP account appropriate in all cases, or should performance on such tasks be reinterpreted as not involving mental state attribution in the way conceived by FP?
The critics certainly argue from considerations of phenomenology that FP accounts have overstated the importance of conscious belief-desire attributions. But the cases they use to make this point and present their alternative conception of social understanding do not obviously cover the case of false belief. For example, they propose that we can perceive the emotions and intentions of others in their expressive behavior—that we do not require theoretical inference or simulative projection from physical behavior to appreciate such states. They also suggest that we often understand people’s behavior in terms of the shared situations and social roles people inhabit, rather than in terms of their beliefs and desires. A favorite example is that we understand a waiter’s actions and can interact with him not because we attribute beliefs and desires to him, but because we understand his social role as a waiter. But unlike their favored cases, a person with false beliefs has an understanding of the world that is not shared with the observer and cannot be perceived in their current behavior. So what should be said on behalf of the critics about false belief understanding?

Let’s start with the online–offline distinction. Surely false-belief understanding comes in not just offline forms, but online forms as well. I can reflect on people’s false beliefs, and even communicate these judgments and behavioral predictions and/or explanations that I can infer from such mental state attributions. But I can also adjust my behavior in light of what I know about people’s false beliefs, what could be called online false-belief understanding. Do the critics permit characterizing such understanding in FP terms—i.e., as the online analog of offline false belief understanding, requiring the application of a concept of BELIEF? Or is this one of Gallagher’s “embodied practices” which does not involve mental state attribution, and involves a different kind of psychological process? To use some of Gallagher’s (2001, 2005) other terminology, is what I’ve been calling online false-belief understanding a form of “mind-
reading”—i.e., an understanding of the “inner”, representational mental states of persons such as beliefs? Or is it a form of “body-reading”—i.e., an “embodied practice” which does not attribute inner mental states to others, but rather attributes some other form of intentionality more intimately tied to behavior?

To make the idea of online false-belief understanding more concrete, let’s consider a few recent experiments in developmental psychology designed to test false belief understanding (Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, in press). Rather than require subjects to verbally or nonverbally make behavioral predictions based on attributions of false beliefs, these experiments test other ways in which false-belief understanding may manifest in children’s online behavior, specifically their looking behavior. Just as in the Sally–Anne version of the standard false belief task, these experiments involve false beliefs created by change in the location of an object.

Onishi and Baillargeon (2005) used a violation-of-expectation paradigm to test whether 15-month-old infants have at least a rudimentary understanding of the false beliefs of other people. After being familiarized with the scene of an agent hiding a toy in one of two locations, then returning later to retrieve the object from the location where she hid the toy, infants were shown scenes where the toy was moved without the adult’s knowledge. Infants were then presented with the adult searching for the hidden toy either (a) where she falsely believed it to be, or (b) where it was actually located, which was contrary to her false belief about its location. Infants reliably looked longer at instances of (b), the so-called “unexpected” event, assuming the child expects the agent to search for the toy where she believes it to be located. This experimental paradigm tests children’s online understanding of others’ false beliefs—i.e., children’s expectations about the behavior of others given what children know about their
epistemic states—rather their ability to verbally or nonverbally report this expectation to a questioner (or even, seemingly, to themselves).

A problem with experiments using average looking time is that they are open to many interpretations about why infants look longer at one condition rather than another. As Southgate et al. (in press) note, the infants in Onishi and Baillargeon’s study might attribute ignorance to the agent rather than a false belief. Thus, infants might look longer at the incongruent event (where the agent’s action is contrary to her false belief) because they do not expect an agent ignorant of an object’s actual location to search for it at that location, rather than because they expect an agent to search for an object in the location she falsely believes it to be located. Southgate et al. attempted to disambiguate these two possibilities by using a predictive looking paradigm, where they measured infants’ anticipatory eye movements prior to seeing an agent searching for a hidden object. Just as in the previous study, infants were first familiarized with videos of an agent watching a toy being hidden in one of two boxes, pausing for a short delay, and then reaching for the toy in that box. A visual and auditory cue was presented during the delay prior to the agent’s reaching for the toy, so the children would learn that the cue signaled that the agent was about to reach for the toy. This was done to increase the likelihood that infants’ first looks to one of the boxes indicated their prediction of where the agent will search for the toy.

In test trials, after the toy was hidden by agent, it was removed from that box while the agent was not looking. This was done to prevent children from being biased in their looking behavior by knowing the actual location of the toy. After the agent returned to looking at the two boxes, the visual and auditory cue was presented during a delay, then the agent reached for the toy in one of the two locations: where she believed it to be located (where she saw it hidden), or
in the other box where she would have no reason to expect it to be hidden. Using eye-tracking technology, the experimenters measured where children first looked after the presentation of the cue—a measure of where the child predicted or expected the agent to search for the toy. During the delay (i.e., prior to the agent reaching for one of the two boxes), these 25-month-olds made their first looks toward, and spent more time looking at, the location in accord with the agent’s false belief. Thus, their looking behavior suggests that the infants expected the agent to look for the toy where she falsely believed it to be located.

So how should we interpret such data? What kind of knowledge about other persons are these infants displaying? One possibility is the traditional FP account. Whether children implicitly simulate the others’ mental states or make use of theoretical knowledge, the standard FP picture is that people possess a concept of BELIEF, and are able to make false-belief ascriptions based on their perception of others’ behavior; such false belief ascriptions are then what cause infants’ predictive looking behavior. For example, one could interpret the infants in the Southgate et al. (in press) study as (implicitly) thinking the following: that person believes the toy is in the box on the left, and he intends to reach for the toy, so I predict he will reach into the box on the left. These children cannot yet articulate this knowledge verbally, but, on this view, the same knowledge about mental states required for the Sally-Anne task is present in these 1.5- to 2-year-old infants. There is no deficit in children’s conceptual knowledge of belief at this young age; there is just some sort of performance deficit such that this knowledge manifests itself in their looking behavior but not their verbal behavior.

The phenomenological critics would surely object to such a FP characterization of the online social understanding displayed by these infants. But what precisely about it would they reject, and why? Even though they claim that what we can directly perceive emotions and
intentions in people’s behavior, and want to treat these states as forms of “bodily intentionality” rather than representational, “inner” mental states, the critics surely cannot treat people’s false beliefs in the same way. What is interesting about false beliefs is that they are not currently perceivable, and thus paradigmatic of why we treat mental states as “inner,” “hidden” and distinct from observable behavior. To understand false beliefs, we must understand that people have points of view on the world which can fail to accord with the actual state of the world. So even if the critics object to the characterization of infants’ looking behavior as driven by conceptual knowledge of belief, how are we supposed to characterize the information driving infants’ looking behavior other than as beliefs, as informational, representational states of persons?

Gallagher (2005) and Ratcliffe (2007, ch. 7) identify one alternative to a representational conception of belief, namely the view that beliefs are “dispositions to act and to experience in various ways” (Gallagher, 2005, p. 214). Both Gallagher and Ratcliffe suggest that possessing a belief does not involve possessing a discrete state, but rather that our belief attributions can be indeterminate and “ambiguous even from the perspective of the believer” (Gallagher, 2005, p. 215). But it is not clear from these authors’ writings what a dispositional theory of belief is an theory of. Is it an account of what beliefs really are? Is it an account of what everyday people take talk about beliefs to be referring to? Is it an account of what we nonverbally represent about other people’s epistemic states for the purpose of nonverbal behavior? Ratcliffe’s (2007, ch. 7) discussion of belief is mostly about the wide range of uses for the term ‘belief’ in our everyday discourse. Gallagher’s (2005) discussion wavers, sometimes referring to what we think and talk about other people in understanding their verbal and nonverbal behavior, and sometimes referring to whether a person “in reality” has a particular belief.
I find such a view fine as an account of our talk about beliefs—I doubt this discourse is as clean and neat as FP’s picture of belief-desire psychology. But I am less satisfied with such an account when attempting to explain psychological processes by which we track people’s epistemic states and act in light of such understanding—i.e., when the focus is on online social understanding such as with the infants in the two experiments I’ve described. In this case, we’re considering a very discrete epistemic state: where an agent thinks a toy is located. Why not treat this as an informational state of that agent, and my understanding of that informational state as representing it? Treating infants as tracking dispositions to behave is just what is claimed by researchers skeptical of attributing mental state understanding to young children. And what exactly is a dispositional understanding of a false belief supposed to look like? How is this distinguished from attributing ignorance to a person? If we restrict ourselves to nonverbal behavior, as we are with our infants, what kinds of behavioral dispositions could distinguish these cases? Without an explicit analysis of false belief from the phenomenological critics, it is difficult to know what to say here.

I contend that it is more parsimonious to claim that infants track false beliefs as informational states of persons, even if we resist that young infants possess the same BELIEF-concept which they do later in development. For instance, it’s unlikely that 2-year-olds understand to the same extent as adults that beliefs are treated by their bearers as true representations of the world which could be either true or false. Young infants may be able to conceive of other people’s representations as differing from what they take the world to be like, but may be unable to conceive of their own beliefs as representations which, even if assumed to be true, can be true or false. Following up on a discussion of Robert Gordon’s version of ST, Jérôme Dokic (2002) makes this point in developing a picture of immature belief-understanding.
that I think is attractive as an explanation of the online false-belief understanding displayed in these experiments. Dokic describes a “non-conceptual view of belief-ascription, according to which the capacity to ascribe beliefs does not require the antecedent possession of the concept of belief as a mental state” (p. 112). On his view, a concept of belief is required to use the result of simulating another’s belief states to have thoughts or make utterances which actually ascribe a belief to a person. But without the concept BELIEF, a person could still use the results of a simulation—i.e., the information about the other’s beliefs gained from pretending to have those beliefs (where ‘pretending’ is not necessarily a conscious or person-level notion)—to drive their behavior, as in the looking behavior of our 1.5-year-olds. Such a person shouldn’t be said to be making unconscious belief-ascriptions, as the person never entertains thoughts using the concept BELIEF, as required by the standard FP account. Yet the person is using information about the person’s mental states gained from a simulation heuristic, and in this way is not representing behavioral dispositions, but actually representing information about the other’s mental states—even though this information can’t be used to make belief-ascriptions using a concept BELIEF, as might be required for success on standard false belief tasks.

4. Conclusion

I do not have the time here to fully develop and defend such an interpretation of the online false-belief understanding displayed by infants in their looking behavior. But I hope to have exposed a limitation in the phenomenological critics’ attack on FP. These experiments of false-belief understanding do support the critics’ contention that online intelligence is very important to social understanding, and that most FP accounts have failed to properly acknowledge this. But the critics have yet to make their case that all online social understanding should be couched in
non-mentalistic terms. While there is reason to restrict to offline cases FP’s analysis of mental state attributions as requiring mental state concepts, it remains to be seen how their non-mentalistic account of intention-reading and situational understanding can be extended to cases of false-belief understanding.

References


